R^3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object

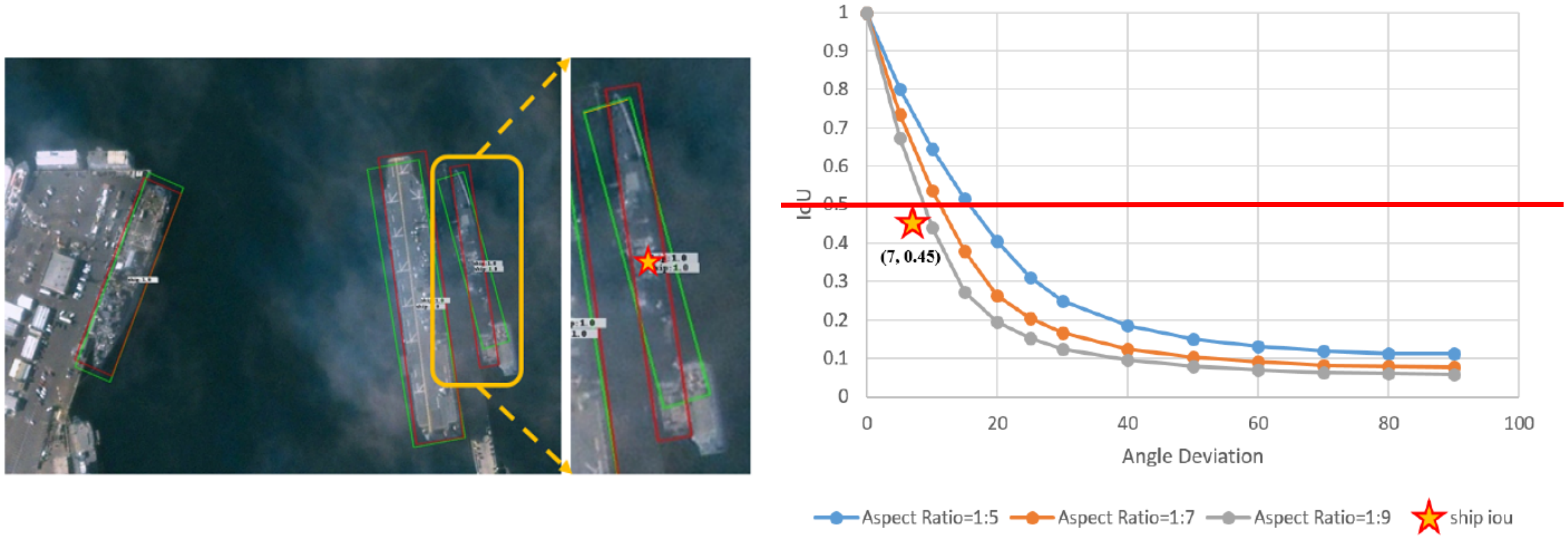
Xue Yang - Shanghai Jiao Tong University

X. Yang, et al. “R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object.” In AAAI21.

Vancouver, Canada. 2021

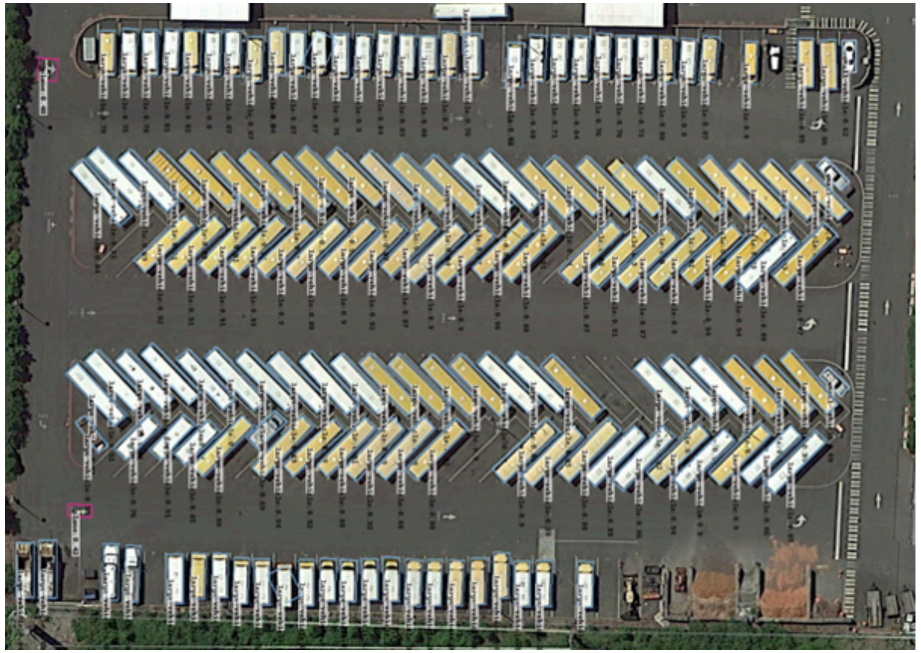
Challenges in Rotation Detection

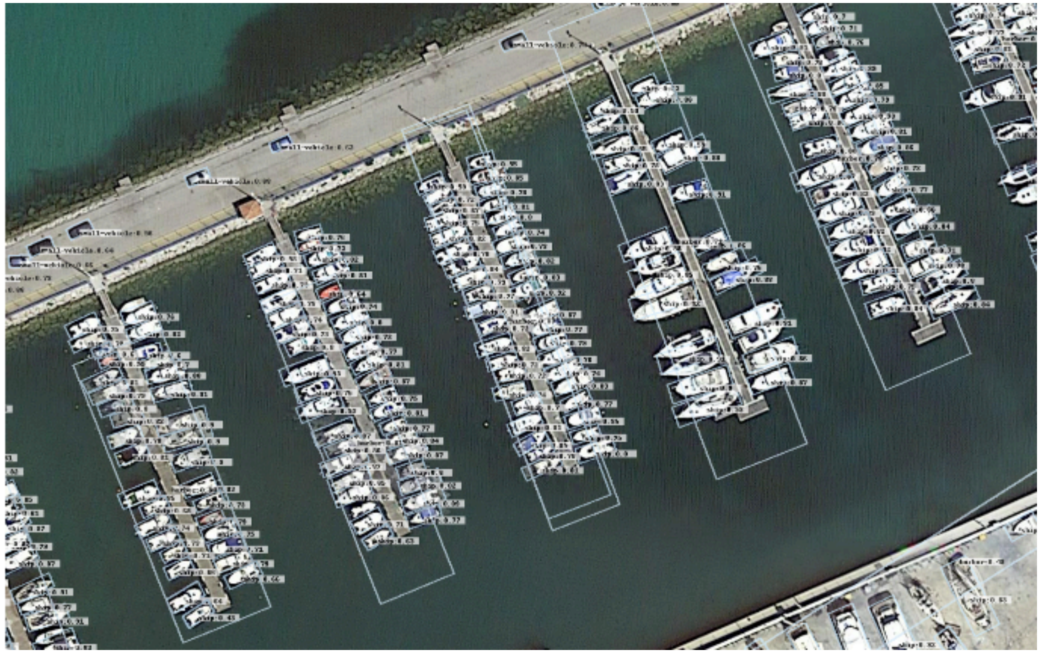
- **Large aspect ratio:** The Skew Intersection over Union (SkewIoU) score between large aspect ratio objects is sensitive to change in angle, as sketched in Figure.



Challenges in Rotation Detection

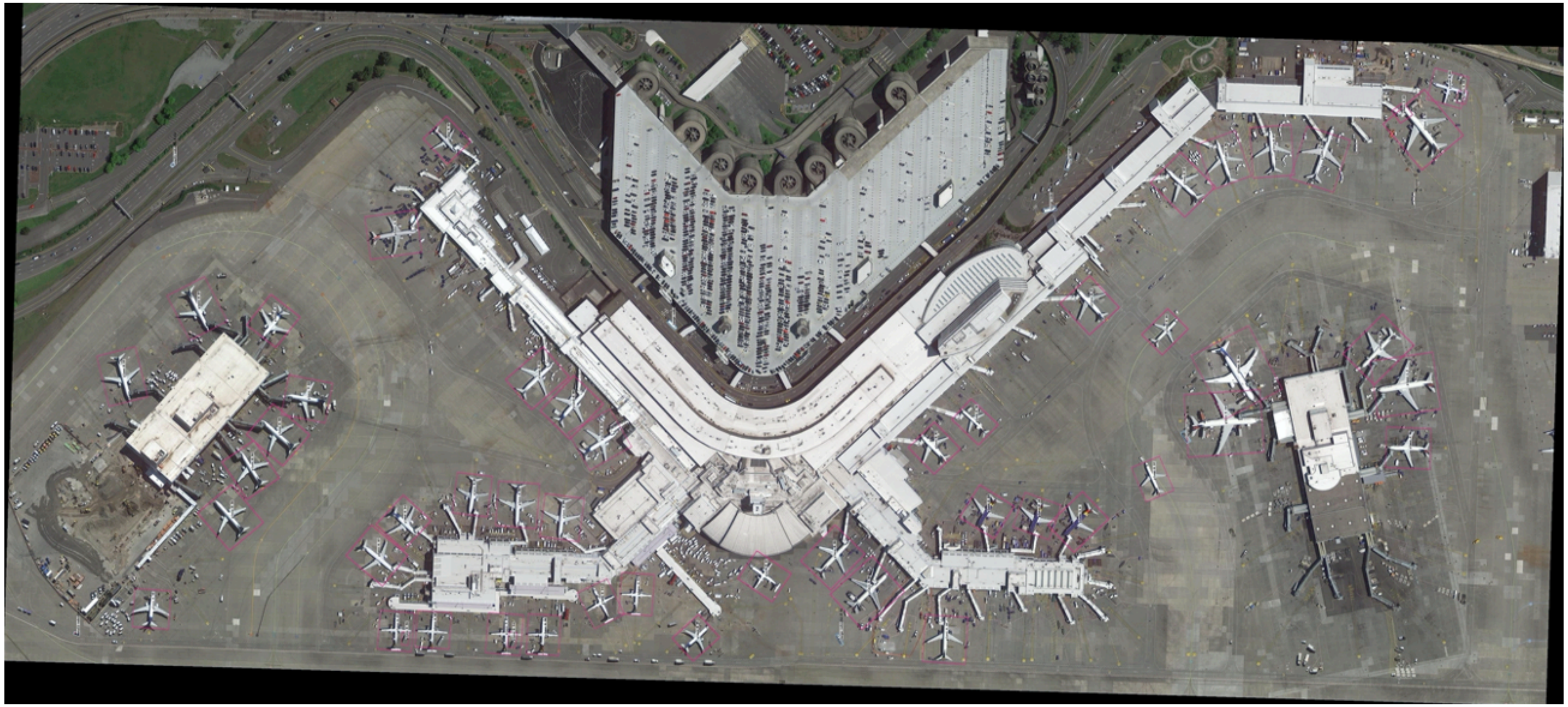
- **Densely arranged:** As illustrated in Figure below, many objects usually appear in densely arranged forms.



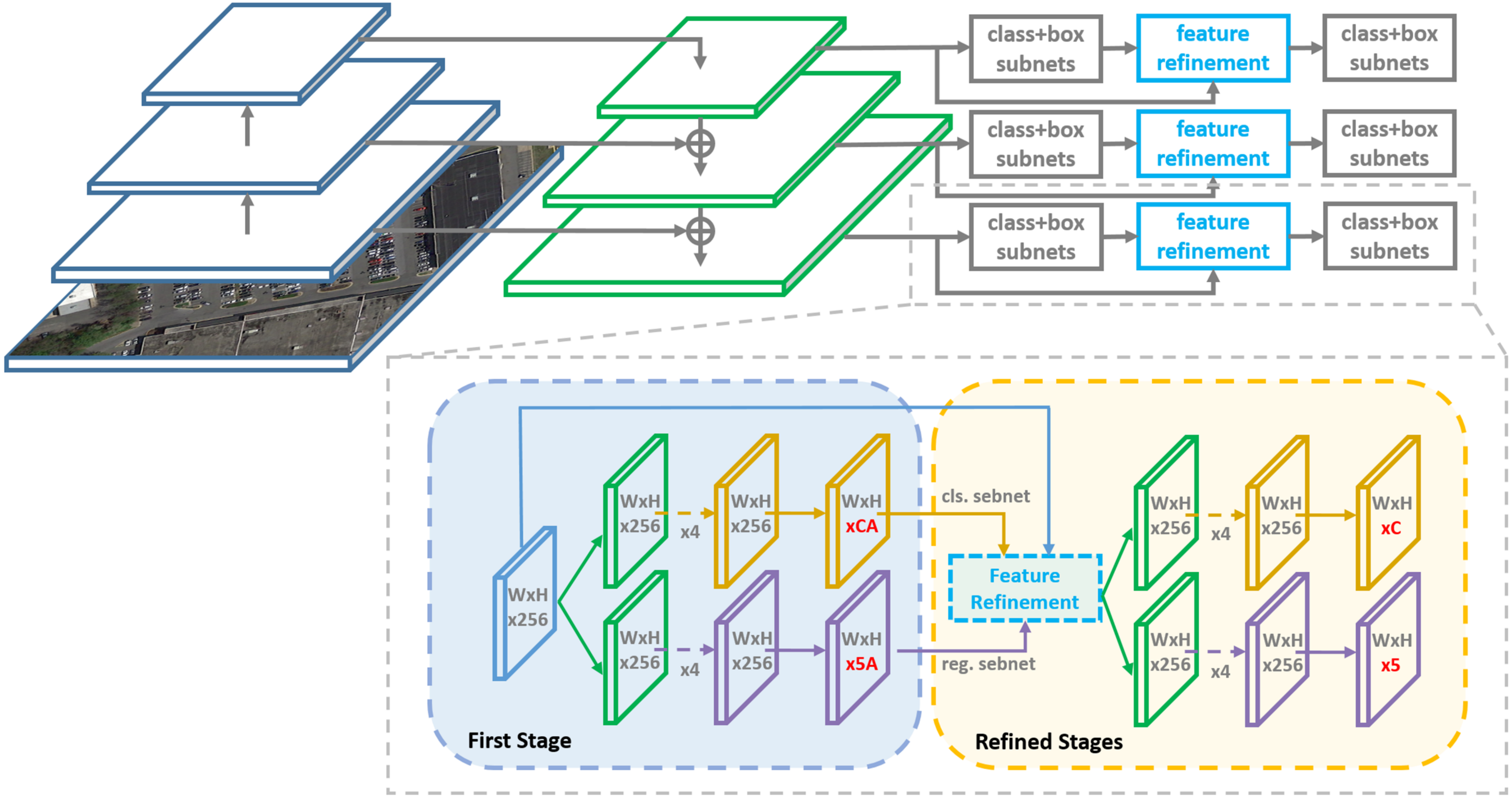


Challenges in Rotation Detection

- **Arbitrary orientations: Objects in images can appear in various orientations, which requires the detector to have accurate direction estimation capabilities.**

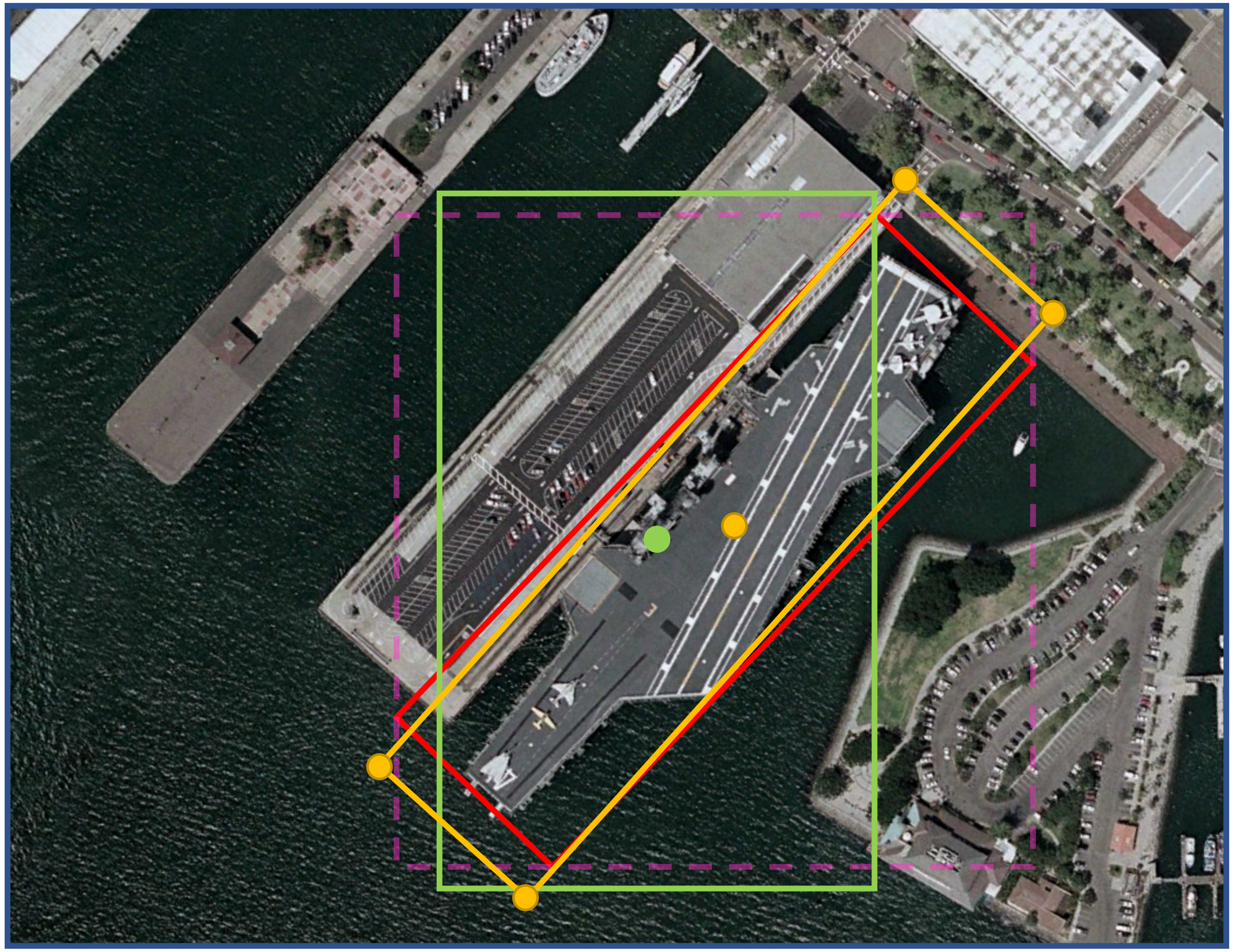


Our Pipeline



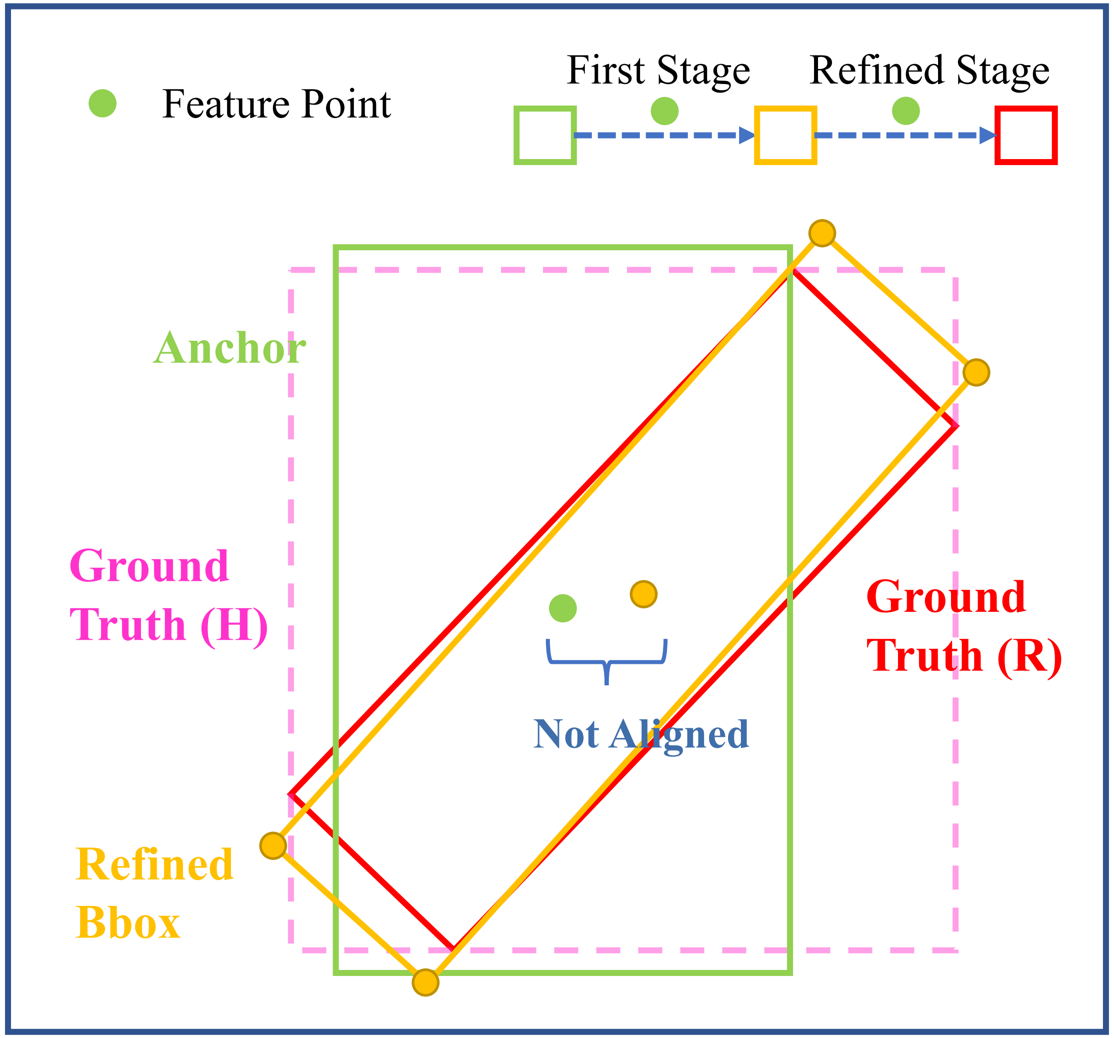
Feature Misalignment

- **Definition:** The current region of interest (RoI) is not aligned with the feature.



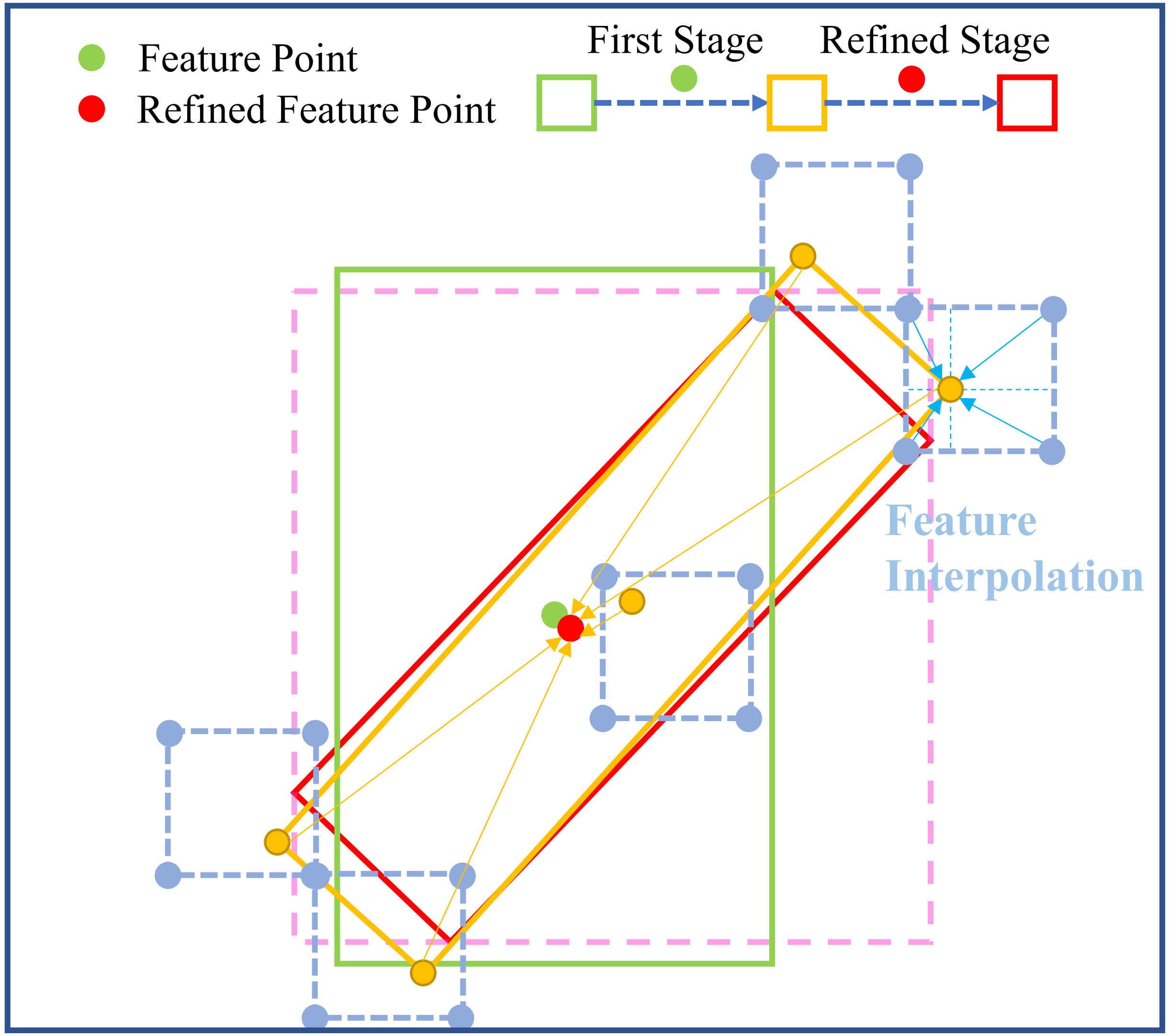
Feature Misalignment

- **Definition:** The current region of interest (RoI) is not aligned with the feature.



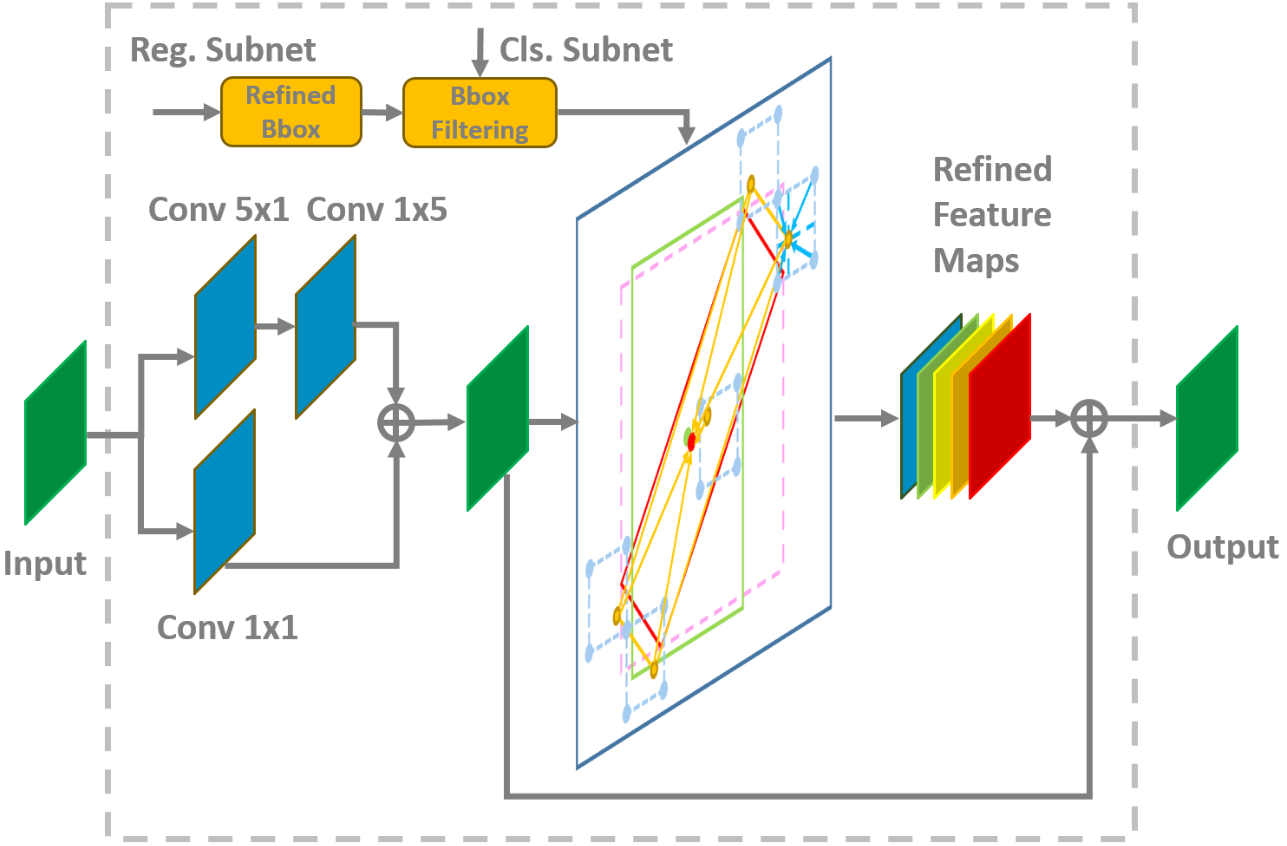
Feature Refinement Module

- **The key idea of FRM is to re-encode the position information of the current refined bounding box to the corresponding feature points through pixel-wise feature interpolation to achieve feature reconstruction and alignment.**



Feature Refinement Module

- The key idea of FRM is to re-encode the position information of the current refined bounding box to the corresponding feature points through pixel-wise feature interpolation to achieve feature reconstruction and alignment.



Feature Refinement

Feature Refinement Module

- **The key idea of FRM is to re-encode the position information of the current refined bounding box to the corresponding feature points through pixel-wise feature interpolation to achieve feature reconstruction and alignment.**

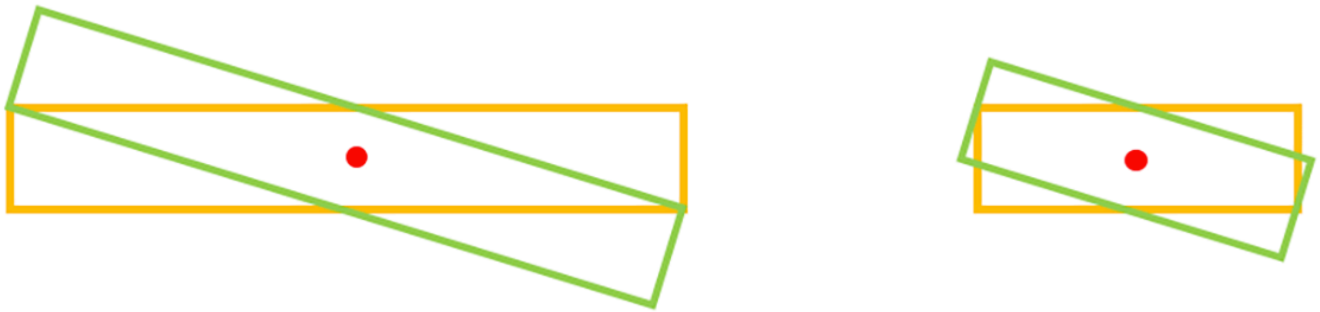
Algorithm 1 Feature Refinement Module

Input: original feature map F, the bounding box (B) and confidence (S) of the previous stage
Output: reconstructed feature map F'

$B' \leftarrow BoxFilter(B, S)$;
$h, w \leftarrow Shape(F)$, $F' \leftarrow ZerosLike(F)$;
$F \leftarrow Conv_{1\times1}(F) + Conv_{1\times5}(Conv_{5\times1}(F))$
for $i \leftarrow 0$ **to** $h-1$ **do**
    for $j \leftarrow 0$ **to** $w-1$ **do**
        $P \leftarrow GetFivePoints(B'(i, j))$;
        for $p \in P$ **do**
            $p_x \leftarrow Min(p_x, w-1)$, $p_x \leftarrow Max(p_x, 0)$;
            $p_y \leftarrow Min(p_y, h-1)$, $p_y \leftarrow Max(p_y, 0)$;
            $F'(i, j) \leftarrow F'(i, j) + BilinearInte(F, p)$;
        end for
    end for
end for
$F' \leftarrow F' + F$;
return F'

Inconsistency between Metric and Loss

- **The evaluation metric of horizontal detection and rotation detection is dominated by Intersection over Union (IoU). However, there is an inconsistency between the metric and regression loss**



$L_{smooth-l1} = 0.274$
$SkewIoU = 0.333$

$L_{smooth-l1} = 0.274$
$SkewIoU = 0.652$

Inconsistency between Metric and Loss

- **The IoU related loss is an effective regression loss function that can solve above problem. However, the SkewIoU calculation function between two rotating boxes is underivable, which means that we cannot directly use the SkewIoU as the regression loss function.**

$$L = \frac{\lambda_1}{N}\sum_{n=1}^{N} obj_n \boxed{\frac{L_{reg}(v_n^{'}, v_n)}{|L_{reg}(v_n^{'}, v_n)|}} \boxed{|f(SkewIoU)|} + \frac{\lambda_2}{N}\sum_{n=1}^{N} L_{cls}(p_n, t_n)$$

$$L_{reg}(v^{'}, v) = L_{smooth-l1}(v_{\theta}^{'}, v_{\theta}) - IoU(v_{\{x,y,w,h\}}^{'}, v_{\{x,y,w,h\}})$$

Experiment

- **Ablative study of each component in our method on the DOTA dataset.**

Method	FRM		approximate	SV	LV	SH	mAP
	BF&FR	LK	SkewIoU loss				
RetinaNet-R				64.64	71.01	68.62	62.76
RetinaNet-H				63.50	50.68	65.93	62.79
R^3Det^*		✓		65.02	67.31	67.31	63.52
R^3Det	✓	✓		65.81	72.76	70.14	66.31
$R^3Det^\dagger$	✓	✓		67.45	73.98	70.27	67.66
$R^3Det^\dagger$	✓	✓	✓	68.04	72.72	76.03	69.50

Experiment

- **Ablation study for number of stages on DOTA.**

#Stages	Test stage	BR	SV	LV	SH	HA	mAP
1	1	39.25	63.50	50.68	65.93	51.93	62.79
2	2	42.72	65.81	72.76	70.14	56.07	66.31
3	3	45.14	67.09	73.70	70.21	56.96	67.29
4	4	44.20	65.30	72.99	70.16	55.70	67.02
3	$\overline{2-3}$	45.08	67.45	73.98	70.27	57.30	67.66

Experiment

- **Experiments with different SkewIoU functions.**

Method	baseline	$-\ln(SkewIoU)$	$1 - SkewIoU$	$\exp(1 - SkewIoU) - 1$
RetinaNet-H	62.79	NAN	65.06 **(+2.27)**	65.34 **(+2.55)**
$R^3Det^\dagger$	67.66	NAN	68.97 **(+2.31)**	69.50 **(+2.84)**

Experiment

- **Comparison with the State-of-the-Art on DOTA.**

	Method	Backbone	MS	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
Two-stage	ICN (Azimi et al. 2018)	ResNet101	✓	81.40	74.30	47.70	70.30	64.90	67.80	70.00	90.80	79.10	78.20	53.60	62.90	67.00	64.20	50.20	68.20
	RADet (Li et al. 2020)	ResNeXt101		79.45	76.99	48.05	65.83	65.46	74.40	68.86	89.70	78.14	74.97	49.92	64.63	66.14	71.58	62.16	69.09
	RoI-Transformer (Ding et al. 2019)	ResNet101	✓	88.64	78.52	43.44	75.92	68.81	73.68	83.59	90.74	77.27	81.46	58.39	53.54	62.83	58.93	47.67	69.56
	CAD-Net (Zhang, Lu, and Zhang 2019)	ResNet101		87.8	82.4	49.4	73.5	71.1	63.5	76.7	**90.9**	79.2	73.3	48.4	60.9	62.0	67.0	62.2	69.9
	Cascade-FF (Hou et al. 2020)	ResNet152		89.9	80.4	51.7	77.4	68.2	75.2	75.6	90.8	78.8	84.4	62.3	64.6	57.7	69.4	50.1	71.8
	SCRDet (Yang et al. 2019b)	ResNet101	✓	89.98	80.65	52.09	68.36	68.36	60.32	72.41	90.85	**87.94**	**86.86**	65.02	66.68	66.25	68.24	65.21	72.61
	FADet (Li et al. 2019)	ResNet101	✓	**90.21**	79.58	45.49	76.41	73.18	68.27	79.56	90.83	83.40	84.68	53.40	65.42	74.17	69.69	64.86	73.28
	Gliding Vertex (Xu et al. 2020)	ResNet101		89.64	85.00	52.26	**77.34**	73.01	73.14	86.82	90.74	79.02	86.81	59.55	**70.91**	72.94	70.86	57.32	75.02
	Mask OBB (Wang et al. 2019)	ResNeXt101	✓	89.56	**85.95**	54.21	72.90	76.52	74.16	85.63	89.85	83.81	86.48	54.89	69.64	73.94	69.06	63.32	75.33
	FFA (Fu et al. 2020)	ResNet101	✓	90.1	82.7	54.2	75.2	71.0	79.9	83.5	90.7	83.9	84.6	61.2	68.0	70.7	76.0	63.7	75.7
	APE (Zhu, Du, and Wu 2020)	ResNeXt101		89.96	83.62	53.42	76.03	74.01	77.16	79.45	90.83	87.15	84.51	**67.72**	60.33	**74.61**	71.84	65.55	75.75
	CenterMap OBB (Wang et al. 2020)	ResNet101	✓	89.83	84.41	**54.60**	70.25	77.66	78.32	87.19	90.66	84.89	85.27	56.46	69.23	74.13	71.56	66.06	76.03
Single-stage	IENet (Lin, Feng, and Guan 2019)	ResNet101	✓	80.20	64.54	39.82	32.07	49.71	65.01	52.58	81.45	44.66	78.51	46.54	56.73	64.40	64.24	36.75	57.14
	PIoU (Chen et al. 2020)	DLA-34		80.9	69.7	24.1	60.2	38.3	64.4	64.8	90.9	77.2	70.4	46.5	37.1	57.1	61.9	64.0	60.5
	P-RSDet (Zhou et al. 2020)	ResNet101	✓	89.02	73.65	47.33	72.03	70.58	73.71	72.76	90.82	80.12	81.32	59.45	57.87	60.79	65.21	52.59	69.82
	O^2-DNet (Wei et al. 2019)	Hourglass104	✓	89.31	82.14	47.33	61.21	71.32	74.03	78.62	90.76	82.23	81.36	60.93	60.17	58.21	66.98	61.03	71.04
	DRN (Pan et al. 2020)	Hourglass104	✓	89.71	82.34	47.22	64.10	76.22	74.43	85.84	90.57	86.18	84.89	57.65	61.93	69.30	69.63	58.48	73.23
	$R^3Det^{\dagger}$ (Ours)	ResNet101		88.76	83.09	50.91	67.27	76.23	80.39	86.72	90.78	84.68	83.24	61.98	61.35	66.91	70.63	53.94	73.79
	R^3Det (Ours)	ResNet152	✓	89.80	83.77	48.11	66.77	**78.76**	**83.27**	**87.84**	90.82	85.38	85.51	65.67	62.68	67.53	**78.56**	**72.62**	**76.47**

Experiment

- **Comparison with the State-of-the-Art on HRSC2016 (left) and UCAS-AOD (right).**

Method	Backbone	Image Size	mAP (07)	mAP (12)	Speed
R^2CNN (Jiang et al. 2017)	ResNet101	800*800	73.07	79.73	5fps
RC1 & RC2 (Liu et al. 2017)	VGG16	–	75.7	–	–
RRPN (Ma et al. 2018)	ResNet101	800*800	79.08	85.64	1.5fps
R^2PN (Zhang et al. 2018b)	VGG16	–	79.6	–	–
RetinaNet-H	ResNet101	800*800	82.89	89.27	14fps
RRD (Liao et al. 2018)	VGG16	384*384	84.3	–	–
RoI-Transformer (Ding et al. 2019)	ResNet101	512*800	86.20	–	6fps
Gliding Vertex (Xu et al. 2020)	ResNet101	–	88.20	–	–
DRN (Pan et al. 2020)	Hourglass104	–	–	92.70	–
SBD (Liu et al. 2019)	ResNet50	–	–	93.70	–
R^3Det*	ResNet101	800*800	89.14	94.98	4fps
RetinaNet-R	ResNet101	800*800	89.18	95.21	8fps
R^3Det	ResNet101	300*300	87.14	93.22	18fps
	ResNet101	600*600	88.97	94.61	15fps
	ResNet101	800*800	**89.26**	**96.01**	12fps
	MobileNetV2	300*300	77.16	84.31	**23fps**
	MobileNetV2	600*600	86.67	92.83	20fps
	MobileNetV2	800*800	88.71	94.45	16fps

Method	mAP	Plane	Car
YOLOv2 (Redmon and Farhadi 2017)	87.90	96.60	79.20
R-DFPN (Yang et al. 2018b)	89.20	95.90	82.50
DRBox (Liu, Pan, and Lei 2017)	89.95	94.90	85.00
S^2ARN (Bao et al. 2019)	94.90	97.60	92.20
RetinaNet-H	95.47	97.34	93.60
ICN (Azimi et al. 2018)	95.67	-	-
FADet (Li et al. 2019)	95.71	**98.69**	92.72
R^3Det	**96.17**	98.20	**94.14**

Thank you!

- Paper: https://arxiv.org/abs/1908.05612
- Code: https://github.com/Thinklab-SJTU/R3Det_Tensorflow

- Contact:
  - Xue Yang: yangxue-2019-sjtu@sjtu.edu.cn
  - Junchi Yan: yanjunchi@sjtu.edu.cn
- Homepage of our lab:
  - http://thinklab.sjtu.edu.cn/